# Accepted Manuscript

Title: Particle Swarm Clustering in clustering ensembles: exploiting pruning and alignment free consensus

Author: José Valente de Oliveira Alexandre Szabo Leandro Nunes de Castro

Please cite this article as: José Valente de Oliveira, Alexandre Szabo, Leandro Nunes de Castro, Particle Swarm Clustering in clustering ensembles: exploiting pruning and alignment free consensus, <![CDATA[Applied Soft Computing Journal]]> (2017), http://dx.doi.org/10.1016/j.asoc.2017.01.035

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# Particle Swarm Clustering in clustering ensembles: exploiting pruning and alignment free consensus

José Valente de Oliveira[1]

*CEOT, University of Algarve, Portugal.*

Alexandre Szabo and Leandro Nunes de Castro

*Natural Computing Laboratory - LCoN, Mackenzie University, S. Paulo, Brazil.*

**Abstract**

A clustering ensemble combines in a consensus function the partitions generated by a set of independent base clusterers. In this study both the employment of particle swarm clustering (PSC) and ensemble pruning (i.e., selective reduction of base partitions) using evolutionary techniques in the design of the consensus function is investigated. In the proposed ensemble, PSC plays two roles. First, it is used as a base clusterer. Second, it is employed in the consensus function; arguably the most challenging element of the ensemble. The proposed consensus function exploits a representation for the base partitions that makes cluster alignment unnecessary, allows for the combination of partitions with different number of clusters, and supports both disjoint and overlapping (fuzzy, probabilistic, and possibilistic) partitions. Results on both synthetic and real-world data sets show that the proposed ensemble can produce statistically significant better partitions, in terms of the validity indices used, than the best base partition available in the ensemble. In general, a small number of selected base partitions (below 20% of the total) yields the best results. Moreover, results produced by the proposed ensemble compare favorably to those of state-of-the-art clustering algorithms, and specially to swarm based clustering ensemble algorithms.

*Keywords:* Clustering, Consensus function, Ensembles, Particle Swarm Clustering, Particle Swarm Optimization, Overlapping partitions.

## 1. Introduction

Clustering plays a pivotal role in disciplines as different as engineering, life sciences, business and economics, defense, psychology, and medicine [1, 2]. Gen-

---